February 2, 2008

Improved Probability Method for Estimating Signal in the Presence of Background

Byron P. Roe[1] and Michael B. Woodroofe[2]

[1] *Department of Physics, University of Michigan, Ann Arbor, MI 48109,*
[2] *Departments of Mathematics and Statistics, University of Michigan, Ann Arbor, MI 48109*

ABSTRACT

A suggestion is made for improving the Feldman-Cousins[1] method of estimating signal counts in the presence of background. The method concentrates on finding essential information about the signal and ignoring extraneous information about background. An appropriate method is found which uses the condition that that the number of background events obtained does not exceed the total number of events obtained. Several alternative approaches are explored.

1. Introduction

Feldman and Cousins,[1] in a recent article, have made major advances towards solving two long-standing problems concerning the use of confidence levels for estimating a parameter from data. The first of these is eliminating the bias that occurs when one decides between using a confidence interval or a confidence bound, after examining the data. The second is finding a confidence interval when the experimental result produces estimators that are close to or past known bounds for the parameters of interest. Feldman and Cousins' method is called the *unified approach* below and is described in Section 2. In the present paper we argue that the unified approach does not make quite enough of an allowance for the known bounds and suggest a modification. The modification is illustrated with the KARMEN 2 Data[2] , where precisely this problem has arisen. The KARMEN group has been searching for a neutrino oscillation signal reported by an LSND experiment.[3] As of Summer 1998, they had expected to see 2.88 ± 0.13 background events and 1.0 - 1.5 signal events, if the LSND results were real, but had seen no events. From their analysis, they claimed to almost exclude the effect claimed by the LSND experiment.

To be specific recall that the Poisson density with mean μ is

$$p_\mu(k) = \frac{1}{k!}\mu^k e^{-\mu} \tag{1}$$

for $k = 0, 1, 2, \cdots$, and let P_μ denote the corresponding distribution function, $P_\mu(k) = p_\mu(0) + \cdots + p_\mu(k)$. Suppose that background radiation is added to a signal producing a total observed count, n say, that follows a Poisson distribution with mean $b + \lambda$. Here the background and signal are assumed to be independent Poisson random variables, with means b and λ respectively. What are appropriate confidence intervals for λ if no events are observed ($n = 0$) or, more generally, if n is smaller than b? For $n = 0$ and a 90% confidence level, the unified intervals all have left endpoints at $\lambda = 0$, while the right endpoints decrease from 2.44 when $b = 0$ to 0.98 when $b = 5$.[♯1] These are the right answers within the formulation of the unified approach.

[♯1] We use here the published numbers for $n = 0$ given by Feldman and Cousins. The numbers we obtain differ slightly. For $b = 3,\ n = 0$, we obtain 0.95 and for $b = 5,\ n = 0$, we obtain 0.77.

The formulation is suspect, however, because the confidence intervals should not depend on b when $n = 0$. For if no events are observed, then both the signal and background radiation must have been zero. It is as if two independent experiments were performed, one for the background and one for the signal. The fact that there were no background events may be interesting but it is not directly relevant to inference about λ once the signal is known, and certainly the a priori expectation b of the background radiation is irrelevant when one knows that the actual background was 0. In this case, the confidence interval for λ should be the same as if one had observed a *signal* of strength 0–either 2.44 using the unified approach, or 2.30 using an upper confidence bound. Statisticians have a name situations like this one. The background radiation is called an *ancillary variable*, because its distribution does not depend on unknown parameters, and conventional statistical wisdom calls for conditioning on ancillary variables when possible.[4] That is what we just did, since conditioning on no background events leaves n as the signal.

Our modification is described in Section 2, where it is compared to the unmodified procedure. For the KARMEN 2 data the modified confidence region is substantially larger than the unmodified one and overlaps the major portion of the LSND region. The modification is compared to a Bayesian solution in Section 4 and shown to agree with it quite well, especially for low counts. Some other possible modifications are discussed briefly in Section 3. Giunti[5] has also proposed a modification of the unified approach and applied it to the KARMEN 2 data. Our approach is contrasted with his in Section 3.

2. An Improved Method

It is not trivial to generalize the method just described to the case of non-zero counts n that may be small compared to the expected background radiation. For if $n > 0$, then it is no longer possible to recover the background and signal. The key to our modification is to remember that a confidence interval consists of values of the parameter that are consistent with the data (that is, are not rejected by an hypothesis test whose significance level is one minus the confidence level). This is also the approach taken by Feldman and Cousins. Suppose, for example, that the expected backgound radiation is $b = 3$ but that only one event is observed ($n = 1$). Is $\lambda = 2$ inconsistent with this observation? From one point of view it is. If $\lambda = 2$, then the probability of observing at most one event is $e^{-5} + 5e^{-5} = 6e^{-5} = .040$, which is less than the usual levels of significance. On the other hand, if only one event is observed, then there can have been at most one background event, and this information should be included in assessing significance. For the probability of at most one background event, $e^{-3} + 3e^{-3} = 4e^{-3} = .199$, is not large, and if the statement $\lambda = 2$ is regarded as an hypothesis, then it seems unfair to include lower than expected background radiation as evidence against it. The way to remove the effect of the low background radiation is to compute the conditional probability of at most one event (total), given at most one background event. The latter is $6e^{-5}/4e^{-3} = 1.5 \times e^{-2} = .203$, which is not less than the usual levels of significance.

Some notation is required to adapt this reasoning to the unified approach. The likelihood function in the signal plus background problem is $L_b(\lambda|n) = p_{b+\lambda}(n)$, where n is the observed count. Following Feldman and Cousins, let $\hat{\lambda} = \max[0, n-b]$ denote the maximum likelihood estimator of λ and let

$$R_b(\lambda, n) = \frac{L_b(\lambda|n)}{L_b(\hat{\lambda}|n)} \tag{2}$$

be the likelihood ratio statistic for testing λ. Then the unified approach consists of taking those λ for which $R(\lambda, n) \geq c(\lambda)$, where $c(\lambda)$ is the largest value of c for which

$$\sum_{k:R_b(\lambda,k)<c} p_{b+\lambda}(k) \leq \alpha \tag{3}$$

and $1-\alpha$ is the desired confidence level. In words, the left side of (3) is the probability that $R_b(\lambda, n) < c$; a level α generalized likelihood ratio test[6] rejects the hypothesis $\lambda = \lambda_0$ if $R_b(\lambda_0, n) < c(\lambda_0)$; and the unified confidence intervals consist of those λ that are not rejected. The modification suggested here consists of replacing $p_{b+\lambda}(k)$ by the conditional probability of exactly k events total given at most n background events. The latter is

$$q^n_{b,\lambda}(k) = \begin{cases} p_{b+\lambda}(k)/P_b(n) & \text{if } k \leq n \\ \sum_{j=0}^{n} p_b(j)p_\lambda(k-j)/P_b(n) & \text{if } k > n, \end{cases} \tag{4}$$

since k total events imply at most n background events when $k \leq n$. Let $\tilde{R}^n_b(\lambda, k)$ denote the likelihood ratio obtained using $q^n_{b,\lambda}(k)$; *i.e.*, $\tilde{R}^n_b(\lambda, k) = q^n_{b,\lambda}(k)/\max_{\lambda'} q^n_{b,\lambda'}(k)$. Let $\tilde{c}_n(\lambda)$ be the largest value of c for which

$$\sum_{k:\tilde{R}^n_b(\lambda,k)<c} q^n_{b,\lambda}(k) \leq \alpha. \tag{5}$$

Then the modified confidence interval consists of those λ for which $\tilde{R}^n_b(\lambda, n) \geq \tilde{c}_n(\lambda)$.

The modified and original unified approaches are compared in Figure 1 for the special case $b = 3$ and $n = 0, \cdots, 15$. Observe that the modified intervals are wider for small n and that there is not much difference for large n. The latter is to be expected, since there is not much difference between $q^n_{b,\lambda}$ and $p_{b+\lambda}$ for large n. In the case of small n, the rationale for the modification is as above. If n is smaller than b, then there was less background radiation than expected, and this information should be used in assessing significance.

For the KARMEN 2 Data, $b = 2.88 \pm 0.13$ and $n = 0$. At the 90% confidence level, the unified approach leads to $0 \leq \lambda \leq 1.08$, and the modified interval leads to $0 \leq \lambda \leq 2.42$. As above, values of λ between 1.08 and 2.42 are found to be inconsistent with the data by the unified approach, but this is due to lower than expected background radiation, and the inconsistency disappears after adjusting for the low background radiation. On the basis of this data, it is not reasonable to exclude the possibility of signal.

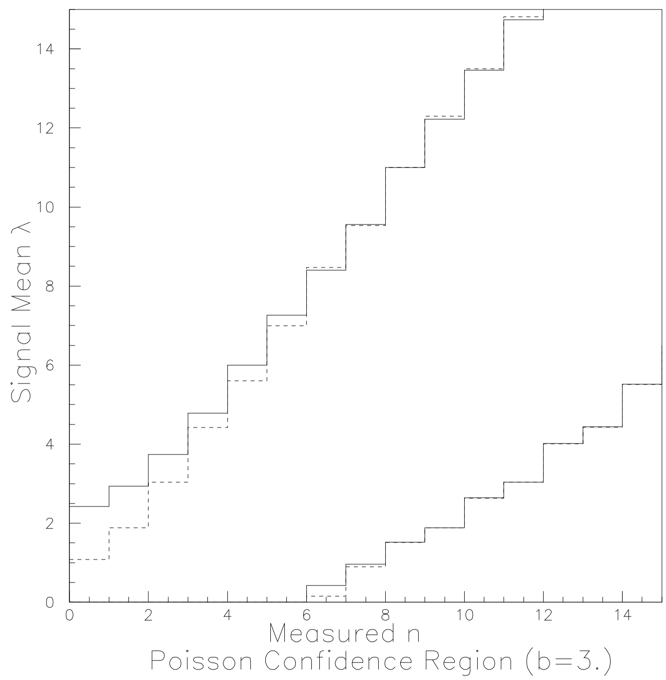

Figure 1. The 90% C.L. region for an unknown Poison signal λ in the presence of a Poisson background $b = 3$. The dashed lines and solid lines correspond to the unified approach and the modified approach, respectively.

To be complete (and fair) Feldman and Cousins were aware of the problem with small counts. For such cases, they suggested reporting the the average upper limit that would be obtained by an ensemble of experiments with the expected background and no true signal, along with the intervals for the observed n. The conceptual difference between our intervals and the unified method is that our confidence levels are conditional and, therefore, refer to a different ensemble. In general terms, the main reason for conditioning is to obtain a model that describes the experiment performed more accurately. The price paid for the more accurate model is often a loss of power, or longer confidence intervals, and the effect can be large, as in the KARMEN data. Of course, power is important, but it is an illusion if the model does not describe the experiment well.

The reader may be familiar with conditioning in the context of contingency tables when some of the row and/or column totals are fixed, have known distributions, or have distributions that only depend on nuisance parameters. In such cases it is appropriate to condition on the known totals, and this affects the distribution of tests statistics and estimators. Fisher's exact test provides a specific example. See Lehmann[7] for a derivation of the exact test and Berkson[8] for criticisms. Other reasons for conditioning arise when the precision with which an experiment was done is observed as part of the outcome. In the

case $n = 0$, our use of conditioning is consistent with these precedents. In the case $n > 0$, however, our use of conditioning goes beyond these established precedents because we condition on an observed bound for an ancillary variable, not the exact value. Our reasons for conditioning, illustrated by the numerical example with $n = 1$ above, are consistent with the precedents. To summarize these reasons: it seems unwise to regard lower than expected background radiation as evidence against a value of λ.

3. Other Possible Modifications

The rationale given for the modification in Section 2 could also have been used to support other modifications. We describe these briefly here and explain our preference for the one described in Section 2. We also contrast our modification with that of Giunti.

The modification described in Section 2 replaces $p_{b+\lambda}$ with $q^n_{b,\lambda}$ in the *derivation* of the unified approach, thus replacing $R_b(\lambda, k)$ by $\tilde{R}^n_b(\lambda, k) = q^n_{b,\lambda}(k)/\max_{\lambda'} q^n_{b,\lambda'}(k)$ in (2) and replacing Equation (3) by Equation (5). An alternative modification would be to keep the unified approach criterion $R_b(\lambda, n)$ but calibrate the associated tests differently, by replacing $p_{b+\lambda}$ with $q^n_{b,\lambda}$ in Equation (3) and, therefore, $c(\lambda)$ with $c_n(\lambda)$ (except that R not $\tilde{R}$ is used). We have explored this approach and found it to be very similar to the one presented. It has the disadvantage that the limits for $n = 0$ are slightly dependent on b.

Our approach may be contrasted with that of Giunti,[5] who has suggested a different modification of the unified approach, called the *new ordering approach*. His physical arguments are along similar lines to ours. However, in detail his approach differs. In the new ordering approach, $R_b(\lambda, n)$ is replaced by $R^{NO}_b(\lambda, n) = p_{\lambda+b}(n)/p_{\lambda^{NO}+b}(n)$ in Equation 3, where λ^{NO} is the Bayes' estimate of λ for a uniform prior. (We shall describe the Bayes' approach further in the next section.) The calibration then proceeds as in Equation 3, using $p_{b+\lambda}(k)$. The resulting intervals are shorter than ours, but depend on b when $n = 0$. Amusingly, our intervals are closer to the Bayesian intervals than are Giunti's intervals, even though our approach is entirely frequentist. See Table 1 below.

In Equation (4),

$$q^n_{b,\lambda}(n) = \frac{p_{b+\lambda}(n)}{P_b(n)} \tag{6}$$

is the conditional probability of n events (total) given at most n background events. This is a very intuitive quantity but, unfortunately, is not a density in n, since $P_b(n) < 1$ for all n and, therefore, $\sum_{n=0}^{\infty} q^n_{b,\lambda}(n) > \sum_{n=0}^{\infty} p_{b+\lambda}(n) = 1$. Of course, $q^n_{b,\lambda}(n)$ could be renormalized by $\kappa(\lambda) := \sum_{n=0}^{\infty} q^n_{b,\lambda}(n)$, and the resulting ratio $q^n_{b,\lambda}(n)/\kappa(\lambda)$ would be a density; but using the ratio in a model would implicitly change the likelihood function. The density then lacks the intuitive appeal of $q^n_{b,\lambda}(k)$, since the definition of the experiment producing this density then becomes unclear.

A closely related quantity is the conditional probability of at most n events total, given at most n background events

$$D_{b,\lambda}(n) = \frac{P_{b+\lambda}(n)}{P_b(n)}. \tag{7}$$

It is not obvious, but $D_{b,\lambda}(n)$ is a distribution function in n for reasons explained below. Let $d_{b,\lambda}(n) = D_{b,\lambda}(n) - D_{b,\lambda}(n-1)$ denote the corresponding density. Still another alternative is to replace $p_{b+\lambda}$ by $d_{b,\lambda}$ in the unified approach. This too led to a procedure that was more complicated and no more efficient than the modification described in Section 2.

To see that $D_{b,\lambda}(n)$ is a distribution function in n, first observe that $\lim_{n\to\infty} D_{b,\lambda}(n) = \lim_{n\to\infty} P_{b+\lambda}(n)/P_b(n) = 1/1 = 1$. So, it suffices to show that $D_{b,\lambda}(n)$ is non-decreasing in n. For this, note that, after some manipulation, $d_{b,\lambda}(n)$ can be written in either of the following forms for $n > 0$:

$$\begin{aligned} d_{b,\lambda}(n) &= d_{0b,\lambda}(n) - \frac{\sum_{k=0}^{n-1} p_{\lambda+b}(k)}{\sum_{j=0}^{n-1} p_b(j)} \times \frac{p_b(n)}{\sum_{i=0}^{n} p_b(i)} \\ &= d_{0b,\lambda}(n) \left[1 - \frac{p_b(n)/\sum_{k=0}^{n-1} p_b(k)}{p_{\lambda+b}(n)/\sum_{j=0}^{n-1} p_{\lambda+b}(j)}\right]. \end{aligned} \tag{8),}$$

where

$$d_{0b,\lambda}(n) = \frac{p_{b+\lambda}(n)}{\sum_{j=0}^{n} p_b(j)} \tag{9).}$$

$D_{b,\lambda}(n)$ will be a non-decreasing function of n if the correction term in the second expression above is always ≤ 1. Using the fact that these are Poisson distributions,

$$\begin{aligned} \frac{p_b(n)/\sum_{k=0}^{n-1} p_b(k)}{p_{\lambda+b}(n)/\sum_{j=0}^{n-1} p_{\lambda+b}(j)} &= \frac{e^{-b}b^n/n!}{\sum_{k=0}^{n-1} e^{-b}b^k/k!} \times \frac{\sum_{j=0}^{n-1} e^{-(b+\lambda)}(b+\lambda)^j/j!}{e^{-(b+\lambda)}(b+\lambda)^n/n!} \\ &= \frac{\sum_{j=0}^{n-1}[1/(b+\lambda)^{n-j}j!]}{\sum_{k=0}^{n-1}[1/(b)^{n-k}k!]} \\ &\leq 1. \end{aligned} \tag{10}$$

The last inequality occurs since $b + \lambda \geq b$.

4. The Bayesian Connection

The discussion in this section makes use of the following identity, which may be established by repeated integrations by parts: if m is any positive integer and $c \geq 0$, then

$$\int_c^\infty p_y(m)dy \equiv \frac{1}{m!}\int_c^\infty y^m e^{-y}dy = \sum_{k=0}^{m} \frac{1}{k!}c^k e^{-c} \equiv P_c(m), \tag{11}$$

This has an amusing consequence: While $q^n_{b,\lambda}(n)$ is not a density in n, it is a density in λ; that is,

$$\int_0^\infty q^n_{b,\lambda}(n)d\lambda = 1. \tag{12}$$

It follows that $q_{b,\lambda}(n)$ is the (formal) posterior distribution that is obtained when λ is given an (improper) uniform distribution over the interval $0 \leq \lambda < \infty$. (It is also the limiting posterior that is obtained if λ is given a (proper) uniform distribution over the interval $0 \leq \lambda \leq \Lambda$ and then Λ is allowed to approach ∞). Moreover, using (11) again, leads to the following curious relation

$$\int_{\lambda_0} q^n_{b,\lambda}(n)d\lambda = D_{b,\lambda_0}(n). \tag{13}$$

That is, the posterior probability that λ exceeds λ_0 given n is the conditional probability of at most n events total given at most n background events when $\lambda = \lambda_0$. Hence, using D, one of our possibilities above, although fully based on a frequentist approach, has some Bayesian justification. Equation (13) also provides frequentist justification for conditioning. For it follows from (13) and Theorem 3.3 of Hwang *et.al.*,[9] that $D_{b,\lambda_0}(n)$ is an admissible p-value for testing $H_0 : \lambda \geq \lambda_0$. Admissibility of the unconditional p-value $P_{\lambda_0}(n)$ is unclear to us at this writing, if $b > 0$.

The Giunti approach, mentioned above, fundamentally uses a partly Bayesian, partly frequentist approach.

Treating $q^n_{b,\lambda}(n)$ as the posterior density in λ leads to Bayesian credible (confidence) intervals of the form $\{\lambda : q^n_{b,\lambda}(n) \geq c_n\}$, where c_n is so chosen to control the posterior probability of coverage; that is,

$$\int_{\{\lambda : q^n_{b,\lambda}(n) \geq c_n\}} q^n_{b,\lambda}(n)d\lambda = 1 - \alpha. \tag{14}$$

Relation (13) is useful in computing the latter integral. The endpoints of these intervals have been computed for selected b and n and are compared to the endpoints of the modified unified approach in the table below.

Table 1. Comparison of Confidence levels for the unified, modified unified, Bayesian, and new ordering approaches described here, for $b = 3$.

	Unified		Modified		Bayesian		New Ord.	
n(observed)	Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper
0	0.0	1.08	0.0	2.42	0.0	2.30	0.0	1.86
1	0.0	1.88	0.0	2.94	0.0	2.84	0.0	2.49
2	0.0	3.04	0.0	3.74	0.0	3.52	0.0	3.60
3	0.0	4.42	0.0	4.78	0.0	4.36	0.0	4.86
4	0.0	5.60	0.0	6.00	0.0	5.34	0.0	5.80
5	0.0	6.99	0.0	7.26	0.0	6.44	0.0	7.21
6	0.15	8.47	0.42	8.40	0.0	7.60	0.28	8.65
7	0.89	9.53	0.96	9.56	0.55	9.18	1.02	9.68
8	1.51	11.0	1.52	11.0	1.20	10.59	1.78	11.2
9	1.88	12.3	1.88	12.22	1.90	11.91	2.49	12.4
10	2.63	13.5	2.64	13.46	2.63	13.19	3.10	13.7

5. Summary

We have suggested a modification to the unified approach of Feldman and Cousins to further improve the estimation of signal counts in the presence of background. It consists of replacing the density function corresponding to the Poisson distribution $p_{b+\lambda}(k)$, with the conditional density function $q^n_{b,\lambda}(k)$. We noted that this method has a clear frequentist justification and is the answer to a clear statistics question.

We compared the results using this modification to the unified approach with the results obtained using the unmodified unified approach. In contradistinction to the old method, the new method leads naturally to sensible results if the observation has fewer events than expected from background events alone.

6. Acknowledgement

We wish to thank Hsiuying Wang for bringing reference [9] to our attention.

REFERENCES

1. G.J. Feldman and R.D. Cousins, Phys. Rev. **D57**, 3873 (1998).
2. B. Zeitnitz *et.al.*, to be published in Prog. Part. Nucl. Physics **40** (1997);
   K. Eitel and B. Zeitnitz for the KARMEN collaboration, Proceedings Contribution to *Neutrinos '98*, Takayama, Japan, June 4-9, 1998.
3. C. Athanassopoulos *et.al.*, Phys. Rev. Lett. **75**, 2650 (1995);
   C. Athanassopoulos *et.al.*, Phys. Rev. Lett. **77**, 3082 (1996);
   C. Athanassopoulos *et.al.*, Phys. Rev. **C54**, 2685 (1996);
   W.C. Louis, representing LSND, Proceedings of the Erice School on Nucl. Physics, 19th course, *Neutrinos in Astro, Particle and Nuclear Physics*, 16–26 September, 1997;
   Response to the FNAL PAC, the BooNE collaboration, April 15, 1998;
   H. White, representing LSND, Neutrinos 98, Takayama, Japan, June 4-9 1998 *Results from LSND*.
4. Reid, N. *The roles of conditioning in inference*, Stat. Sci. **10**, No. 2, 139-199 (1995).
5. C. Giunti, *Statistical interpretations of the null result of the KARMEN 2 experiment*, DFTT 50/98 (hep-ph/9808405), August 25, 1998.
6. J. Rice, *Mathematical Statistics and Data Analysis*, 2nd ed., Duxbury (1995).
7. E.L. Lehmann, *Testing Statistical Hypotheses*, 2nd ed., pp. 156-162, Wadsworth (1986).
8. J. Berkson, *In dispraise of the exact test,* J. Statist. Plan. Inf. **2**, 27-40 (1978).
9. J.T. Hwang, G. Casella, C. Robert, M. Wells, and R. Farrell, *Estimation of accuracy in testing,* Ann. Statist., **20**, 490-509 (1992).